Data Fusion Using Source Separation:

Why and How to Account for Multiple Types of Statistical Diversity

Tülay Adalı
University of Maryland Baltimore County
Machine Learning for Signal Processing Lab
Baltimore, MD
http://mlsp.umbc.edu/

The 20th Annual Signal & Image Sciences Workshop
CASIS, Lawrence Livermore National Laboratory
Livermore, CA

May 18, 2016

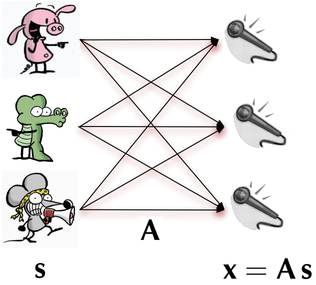
What is diversity?

Of differing elements or qualities

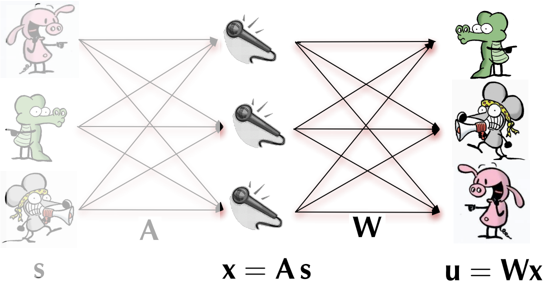
Certain possibilities might be more exciting than others

A statistical property that enables identification, and for multiple datasets, also links them...

Independent component analysis (ICA) assumes that the underlying sources are statistically *independent*



Independent component analysis (ICA) assumes that the underlying sources are statistically *independent*



Independence is a strong assumption and hence enables a solution subject to *only* a permutation and scaling ambiguity

Independence is also a plausible assumption in many applications

ICA is based on the classical cocktail party problem

Artist: Annie Campbell

But besides the cocktail party problem (audio and speech processing) applications of ICA include

- medical data analysis and fusion (e.g., fMRI, EEG, ECG, sMRI)
- noise/interference removal
- communications (e.g., multiuser detection in CDMA)
- data mining
- sensor array processing
- remote sensing
- financial and other time series analysis
- feature extraction for detection/classification

Outline

Very brief history of ICA

- Hérault and Jutten, Snowbird, Utah, 1986
  Source separation with nonlinear decorrelations
- Jutten, Hérault, and Guerin, 1988
  An adaptive algorithm
- Comon, Cardoso, early 1990s
  ICA, cost/contast functions
  Explicit computation of higher-order statistics, e.g., JADE
- Bell and Sejnowski, 1995
  Information maximization: Infomax
- Hyvärinen 1997, 1999
  Maximization of non-Gaussianity: FastICA
- First International Workshop on ICA and Signal Separation, 1999, Aussois, France
- 13*th* LVA/ICA Conference will be held in Grenoble, France

Multiple "routes" for ICA

Given that the sources s_n in $\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix}$ are mutually independent,

different types of *diversity*—statistical property—can be used to achieve ICA:

- Non-Gaussianity (HOS)
  Infomax, FastICA, EFICA, JADE, EBM, RADICAL, and many others
- Sample (linear) dependence (nonwhiteness)
  AMUSE, SOBI, WASOBI, and others
- Nonstationarity
- Noncircularity—*for complex data*

Why not account for multiple types of diversity jointly?

Then, use *random processes* to define the latent model

- Sample index
- Unknown source (component) vector
- Unknown invertible mixing matrix
- Mixtures

$$\mathbf{x}(v) = \mathbf{A}\,\mathbf{s}(v), \quad v = 1, \ldots, V$$

Given $\mathbf{x}(v) = \mathbf{A}\,\mathbf{s}(v)$ where $\mathbf{x}(v), \mathbf{s}(v) \in \mathbb{R}^N$ and $\mathbf{A} \in \mathbb{R}^{N \times N}$

Estimate a demixing matrix $\mathbf{W}$, such that the source estimates are given by

$$\mathbf{u}(v) = \mathbf{W}\mathbf{x}(v)$$

Mutual information *rate* can account for *all four types of statistical diversity*

We can estimate $\mathbf{u}(v) = \mathbf{W}\mathbf{x}(v)$ where $u_n(v) = \mathbf{w}_n^T \mathbf{x}(v)$ by

$$\mathcal{I}_r(\mathbf{W}) = \sum_{n=1}^{N} H_r(u_n) - \underbrace{H_r(\mathbf{u})}_{\log|\det \mathbf{W}| + H_r(\mathbf{x})}$$

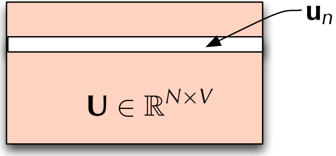
where the entropy rate $H_r(u_n) = \lim_{v \to \infty} \left[H\left[u_n(1), \ldots, u_n(v)\right]/v\right]$ and $H(u_n) = -E\{\log p_{s_n}(u)\}$

Hence, can achieve ICA by minimizing

$$\mathcal{I}_r(\mathbf{W}) = \sum_{n=1}^{N} H_r(u_n) - \log|\det \mathbf{W}| - C$$

For given $\mathbf{X} \in \mathbb{R}^{N \times V}$, can write the likelihood function

Define $\mathbf{u}_n$ as the nth row of $\mathbf{U} = \mathbf{W}\mathbf{X}$



$$\mathcal{L}_{\mathrm{ICA}}(\mathbf{W}) = \sum_{n=1}^{N} \log p_{s_n}(\mathbf{u}_n) + V \log |\det \mathbf{W}|$$

Note that optimality properties of maximum likelihood imply the estimation of both $\mathbf{W}$ *and* $p_{s_n}(\cdot)$

Identifiability of the ICA model

- Compute the Hessian of $\mathcal{L}_{\text{ICA}}(\mathbf{W})$ with respect to the global demixing matrix $\mathbf{G} = \mathbf{WA}$, and evaluate at the optimum $\mathbf{G} = \mathbf{I}$, hence $u_n = s_n$
- The Fisher information matrix can be characterized by the 2×2 matrix, *i.e.*, by pairwise relationship of sources for $1 \leq m < n \leq N$

$$\mathbf{J}_{m,n} = \begin{bmatrix} \kappa_{m,n} & 1 \\ 1 & \kappa_{n,m} \end{bmatrix}, \quad \text{where} \quad \kappa_{n,m} = \text{trace}\left(E\left\{ \psi(\mathbf{s}_n)\psi^\top(\mathbf{s}_n) \right\} \mathbf{R}_m \right),$$

$$\psi(\mathbf{s}_n) = -\frac{\partial \log p_{\mathbf{s}_n}(\mathbf{s}_n)}{\partial \mathbf{s}_n} \in \mathbb{R}^V, \quad \text{and} \quad \mathbf{R}_n = E\{\mathbf{s}_n\mathbf{s}_n^\top\} \in \mathbb{R}^{V \times V}$$

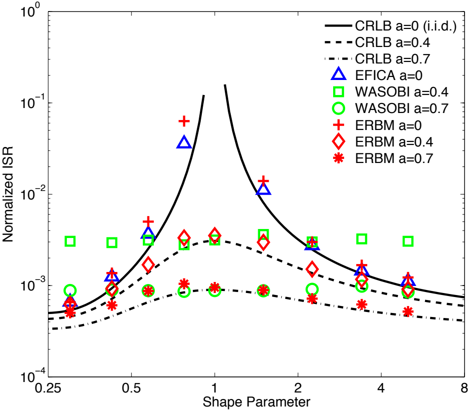
Separation is possible—$\mathbf{J}_{m,n}$ remains positive definite—as long as sources are not both Gaussian with proportional covariance matrices $\mathbf{R}_m = \delta^2 \mathbf{R}_n$

Since for i.i.d. sources, $\sigma_m^2 = \delta^2 \sigma_n^2$, in this case, can identify a single Gaussian

[Cardoso, 2010]

Performance improves with the addition of each type of diversity

(Induced) CRLB as a function of shape parameter β



Two sources:

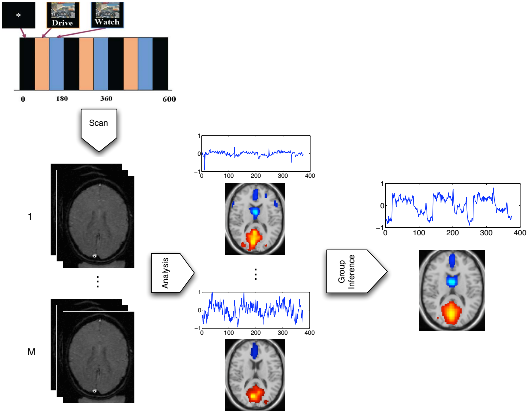
$s_1(v)$ is i.i.d. & GGD with
β – non-Gaussianity

$s_2(v) = as_2(v-1) + \xi(v)$
with i.i.d. and Gaussian $\xi(v)$
a – sample dependence

- (Induced) CRLB
- EFICA – only non-Gaussianity
- WASOBI – only sample dependence
- ERBM – *Both non-Gaussianity and sample dependence*

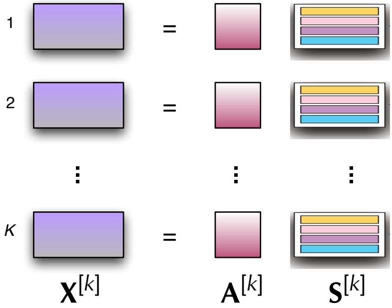
Two Gaussians ($\beta = 1$) are identifiable as long as one is not i.i.d.

Joint analysis of multiple datasets arises in many applications



For example in fMRI experiments, we are interested in group inferences or studying differences among groups

Other examples



- Remote Sensing:
  Hyperspectral data fusion and analysis, beamforming problems
- Video/image Processing:
  Object detection, scene analysis
- Medical Image Analysis:
  Group fMRI, EEG over multiple epochs, or multiple subjects
- Medical Data Fusion:
  Fusion of fMRI, sMRI, EEG, and genetic, or multi-task data
- Audio/Speech Processing:
  Frequency domain ICA

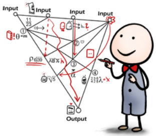
Note statistical dependence as the source of diversity across the datasets

Why perform a *joint* analysis?

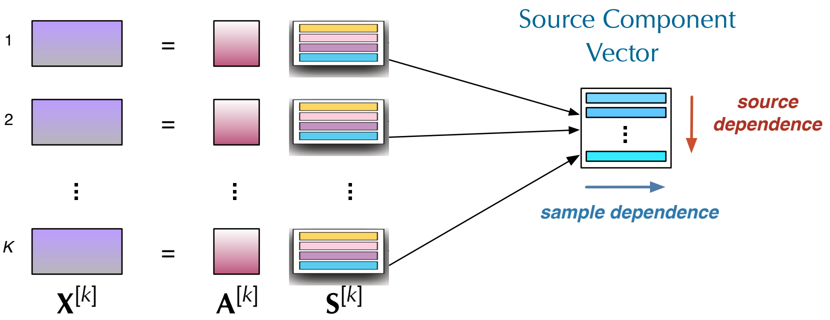
- When performing analysis of multiple datasets, we can perform ICA individually on each dataset
- However, resulting estimates have *different permutations* making multi-dataset analysis difficult

By performing a joint analysis, can resolve the permutation ambiguity across datasets

More importantly, can make use of the true multivariate nature of the data and the statistical dependence across the datasets for better performance



Now, can take advantage of *one more statistical diversity*



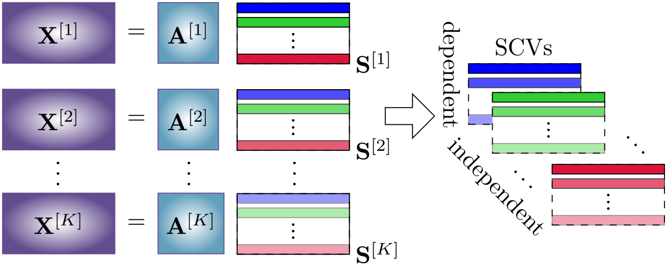
And of course ***HOS, nonstationarity, and noncircularity*** *as well*

Multi-dataset analysis – *Joint blind source separation*

Original ICA: $\mathbf{X} = \mathbf{A}\mathbf{S}$ $\quad \mathbf{U} = \mathbf{W}\mathbf{X}, \quad \mathbf{X}, \mathbf{U}, \mathbf{S} \in \mathbb{R}^{N\times V}$

Joint analysis: $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]} \quad \mathbf{U}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}, \quad k = 1, \ldots, K$

$$\begin{bmatrix} \mathbf{X}^{[1]} \\ \vdots \\ \mathbf{X}^{[K]} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{[1]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{[K]} \end{bmatrix} \begin{bmatrix} \mathbf{S}^{[1]} \\ \vdots \\ \mathbf{S}^{[K]} \end{bmatrix} \Longleftrightarrow \mathbf{X} = \mathbf{A}\mathbf{S} \text{ where } \mathbf{A} = \oplus \sum_{k=1}^{K} \mathbf{A}^{[k]}$$



Independent vector analysis (IVA) is also achieved by minimizing mutual information rate

Minimize MI rate for N sources and K datasets

$$\mathcal{I}_r^{\text{IVA}}(\mathcal{W}) = \sum_{n=1}^{N} H_r\left(\mathbf{u}_n\right) - \sum_{k=1}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - C$$

now, among source component vectors (SCVs), $\mathbf{s}_n(v)$ estimated by $\mathbf{u} = \mathbf{W}\mathbf{x}$

Can rewrite the cost as

$$\mathcal{I}_r^{\text{IVA}}(\mathcal{W}) = \sum_{n=1}^{N}\left(\sum_{k=1}^{K} \underbrace{H_r[u_n^{[k]}]}_{\text{Entropy rate}} - \underbrace{\mathcal{I}_r[\mathbf{u}_n]}_{\text{SCV MI rate}}\right) - \sum_{k=1}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - C$$

[Kim, *et al.*, 2006; Anderson, *et al.*, 2014]

We can similarly write the log likelihood

For given $\mathbf{X}^{[k]}, k = 1, \ldots, K$, we have

$$\mathcal{L}_{\text{IVA}}(\mathcal{W}) = \sum_{n=1}^{N} \log\left(p_n(\mathbf{U}_n)\right) + V \sum_{k=1}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right|$$

where the score function for $\mathbf{U}_n$ is

$$\boldsymbol{\Psi}_n\left(\mathbf{U}_n\right) = -\frac{\partial \log\left(p_n\left(\mathbf{U}_n\right)\right)}{\partial \mathbf{U}_n} \in \mathbb{R}^{K \times V}$$

and the source component matrix (SCM) $\mathbf{U}_n$ is $\mathbf{u}_n(v)$ for $v = 1, \ldots, V$

Identification conditions for IVA are determined similarly

- Compute the Hessian of $\mathcal{L}_{\text{IVA}}(\mathcal{W})$ wrt the block diagonal global demixing matrix $\mathbf{G} = \oplus_{k=1}^{K} \mathbf{W}_k \mathbf{A}_k = \oplus_{k=1}^{K} \mathbf{G}_k$ at the optimum $\mathbf{G} = \mathbf{I}$, hence $\mathbf{U}_n = \mathbf{S}_n$
- Fisher information matrix is *now* characterized through interactions of *two block matrices*

$$\mathbf{J}_{m,n} \triangleq \begin{bmatrix} \mathcal{K}_{m,n} & \mathbf{I}_K \\ \mathbf{I}_K & \mathcal{K}_{n,m} \end{bmatrix} \in \mathbb{R}^{2K \times 2K}, \;\; 1 \leq m < n \leq N$$

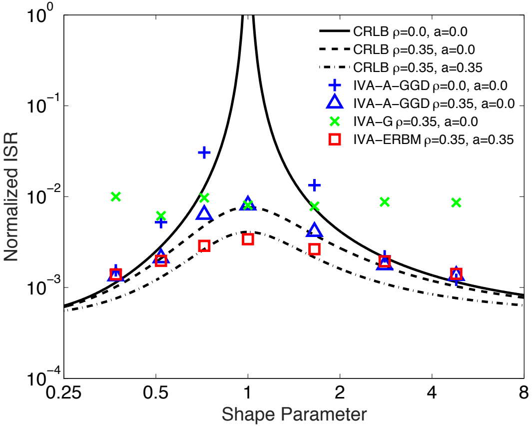
where $\{\boldsymbol{\mathcal{K}}_{m,n}\}_{k_1,k_2} = \frac{1}{V} E\left\{ \left(\boldsymbol{\psi}_m^{[k_1]}\right)^{\mathsf{T}} \mathbf{s}_n^{[k_1]} \left(\mathbf{s}_n^{[k_2]}\right)^{\mathsf{T}} \boldsymbol{\psi}_m^{[k_2]} \right\}$ and $\boldsymbol{\Psi}_n(\mathbf{S}_n) = -\frac{\partial \log\left(p_n\left(\mathbf{S}_n\right)\right)}{\partial \mathbf{S}_n}$, $\boldsymbol{\psi}_n^{[k]} = \boldsymbol{\Psi}_n^{\mathsf{T}} \mathbf{e}_k$

Identification is possible as long as no two SCMs have α-Gaussian components for which $\mathbf{R}_m = (\mathbf{I}_V \otimes \mathbf{D})\, \mathbf{R}_n\, (\mathbf{I}_V \otimes \mathbf{D})$, *for* $1 \leq m \neq n \leq N$

[Anderson, *et al.*, 2014]

Performance *again* improves with the addition of each type of diversity

(Induced) CRLB as a function of shape parameter β



Two sets of sources (SCVs):

- Multivar. i.i.d. GGD pair with β and ρ
- First-order AR vector process with $\mathbf{A} = a\mathbf{I}$

- (Induced) CRLB
- IVA-A-GGD – HOS

  IVA-A-GGD – HOS & source dependence
- IVA-G – source correlation
- IVA-ERBM – *HOS & source & sample dependence*

Now, can identify i.i.d. (real) Gaussians as well...

Short intermediate summary—*Focus on diversity*

*Using multiple types of diversity, **jointly**, we*

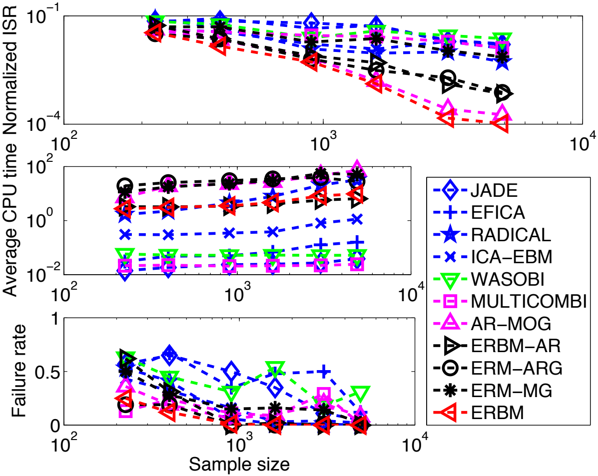
- can identify a broader class of signals, and
- maximally use all available information, design *efficient* estimators

both for single and multi-set data analysis/fusion

Need to incorporate diversity into solutions, through

- explicit modeling—*single and multiple datasets*
- identifying *relevant* sources of diversity to define the datasets and modify the model—*multiple datasets*

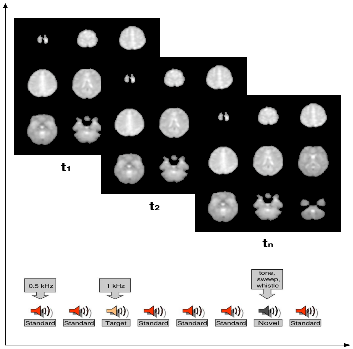
A classic example: “Artificial mixture” of images



Separation performance for artificial mixture of eight images

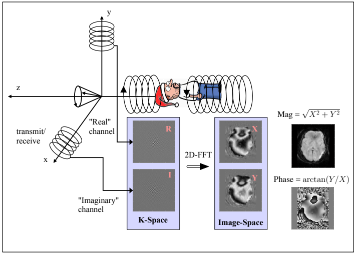
What about when applied to practical problems?

Functional MRI analysis—A fruitful application domain for ICA



Functional MRI (fMRI) reports on local brain hemodynamics

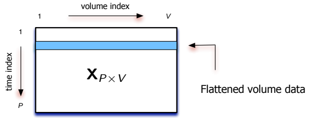
MRI signal is acquired as a quadrature signal using two orthogonal detectors



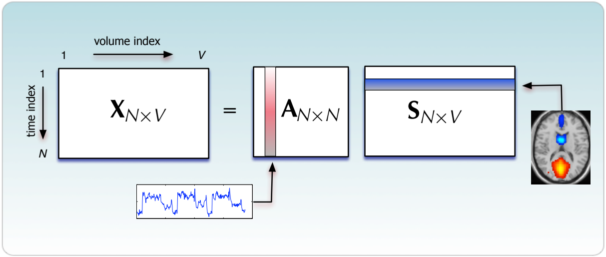
Hence, it is inherently complex valued

Spatial ICA of fMRI finds maximally spatially independent components (spatial maps)

Form the observation (mixture) matrix **X** by stacking volume data at each time instant

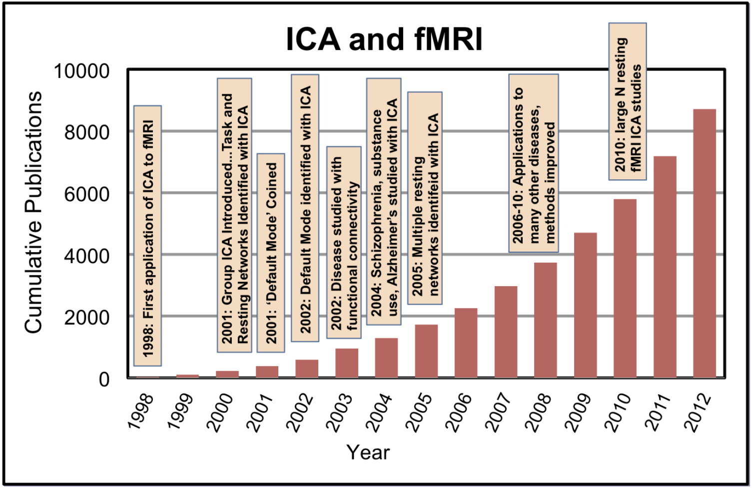


Spatial ICA of fMRI



[McKeown, *et al.*, '98]

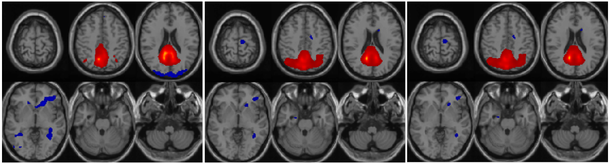
ICA has been widely used for fMRI analysis



[Calhoun and Adalı, 2012]

Flexible ICA algorithms such as EBM and ERBM provide better performance

ICA estimates of default mode network (DMN)

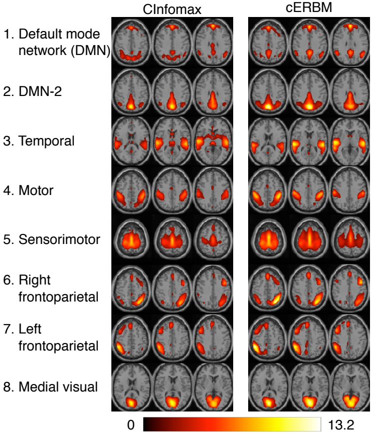


Infomax EBM ERBM

	Infomax	EBM	ERBM
Number of voxels overlapping with the mask	2386	3291	3328
Sensitivity of t map with corresponding mask	0.73	0.82	0.82

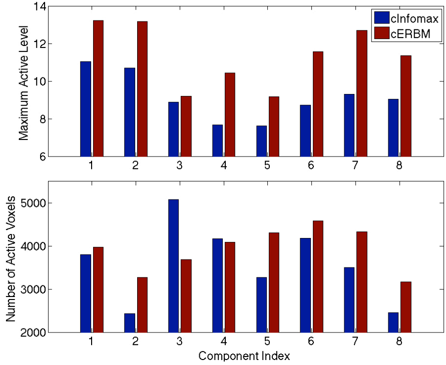
[Du, *et al.*, 2011]

Incorporation of multiple types of diversity yields better estimates



Z-maps of fMRI data from 100 subjects

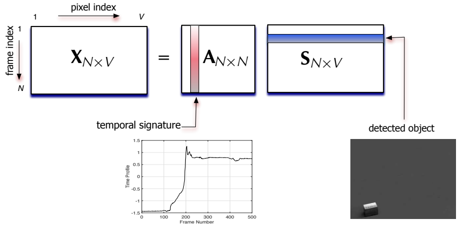
Maximum level of activation and the number of active voxels



[Du *et al.*, 2014 and 2016]

ERBM provides better performance for a video application as well

Abondoned object detection

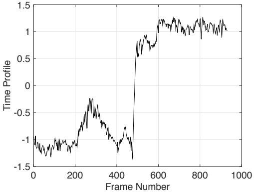


Sample video

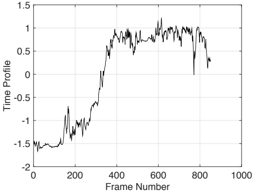
Summarize temporal dynamics through independent components

Other detection examples—*videos of varying difficulty*

Parked car on a main road

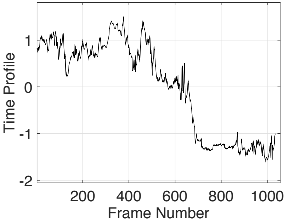


A suitcase left in a London metro station



For a night video, detection is possible only using sample dependence & HOS—*using ERBM*

Detected component and temporal signature using ERBM



Infomax, EFICA, EBM, and WASOBI all fail to detect the component

Original frame

Multi-set vs multi-modal fusion

- Multi-set data

  Information collected using the same *modality* at different conditions, observation times, using multiple experiments or subjects,. . .

  Datasets are of the same type and dimension

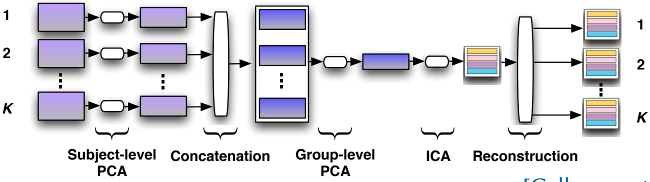
- Multi-modal data

  Information collected through different types of detectors/sensors
  - Medical imaging data, hemodynamic response and electrical activity
  - Remote sensing, optical, and radar imagery
  - Audio and video data

  Datasets are of different nature, resolution, and size

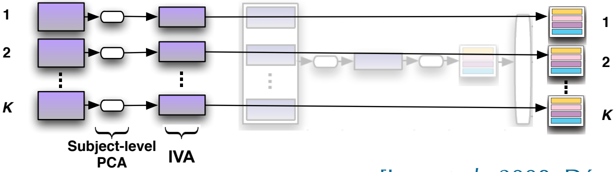
Group ICA vs IVA for the analysis of multi-subject fMRI data

Group ICA defines a group subspace and performs a single ICA



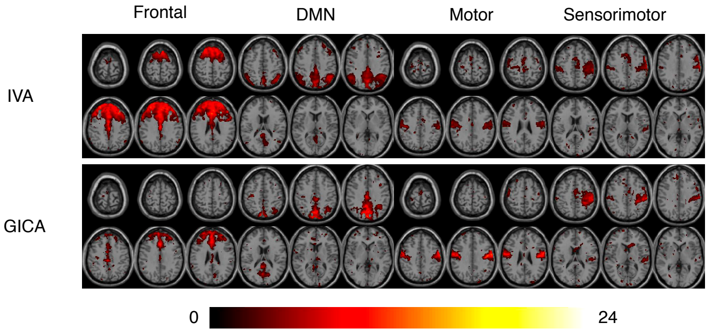
[Calhoun, *et al.*, 2001]

IVA avoids the common subspace, hence is expected to better preserve variability



[Lee, *et al.*, 2008, Déa, *et al.*, 2011]

IVA leads to better performance in real fMRI data



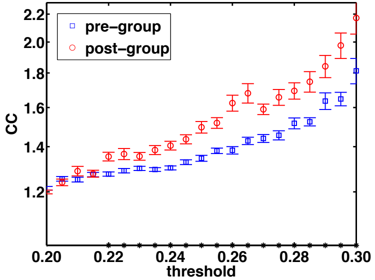
Thresholded *t*-maps at a significance level of 0.05 for stroke patients performing a motor task

[Laney, *et al.*, 2015]

And results in lower *p*-values

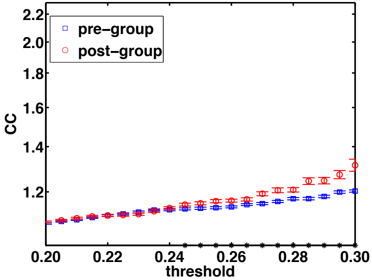
Average clustering coefficient using graph-theoretical analysis

IVA



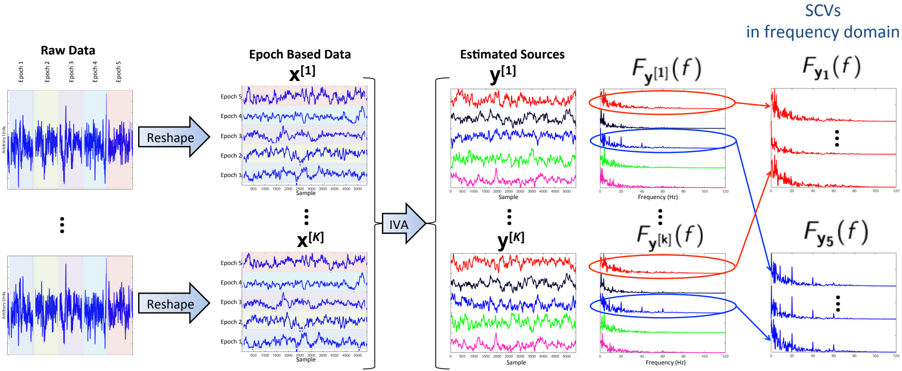
$p < .0088$

GICA



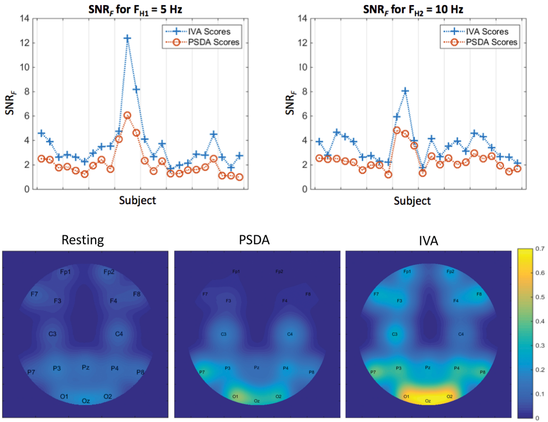
$p < .032$

IVA for enhancement of steady state visually evoked potentials (SSVEP)



[Emge, *et al.*, 2015]

IVA consistently shows enhancement across subjects and experiments for SSVEP



SNR$_F$ scores for one subject with rest state as reference

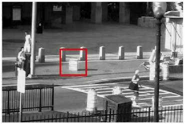
[Emge, *et al.*, 2016]

IVA provides advantages for abandoned object detection

Use IVA to perform *joint separation* of video from red, green, and blue channels

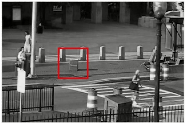
RGB image

Monochrome image



Red channel

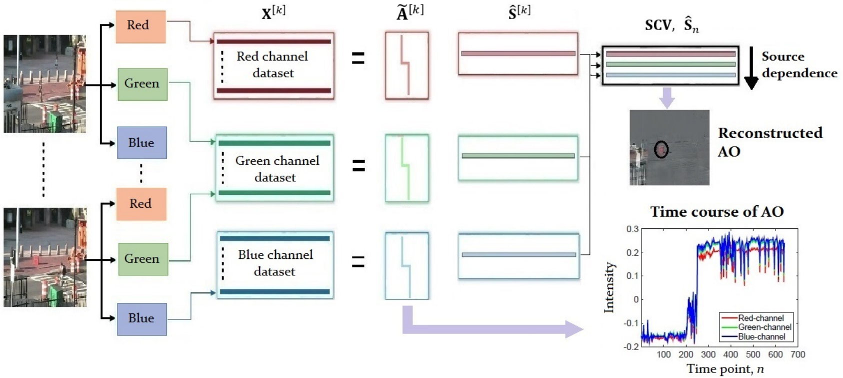
Green channel



Blue channel

- Better estimation due to use of additional diversity, *source dependence*
- Additional robustness for detection through use of three temporal sequences
- Potential to make use of color information

IVA improves the detection power



t-statistics for the step response

Video	ICA-GGD	IVA-GGD		
Abandoned Box	123.13	108.49	136.30	134.44
Tramstop	99.95	124.63	119.55	117.91
PV-Easy	287.15	92.42	90.13	91.27
PV-Hard	55.01	77.81	71.98	73.78
PV-Night	46.84	58.23	59.89	59.23

Multi-set vs multi-modal fusion

- Multi-set data

  Information collected using the same *modality* at different conditions, observation times, using multiple experiments or subjects,. . .

  Datasets are of the same type and dimension

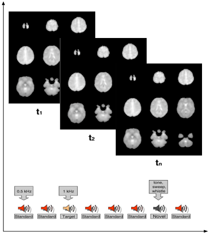
- Multi-modal data

  Information collected through different types of detectors/sensors
  - Medical imaging data, hemodynamic response and electrical activity
  - Remote sensing, optical, and radar imagery
  - Audio and video data

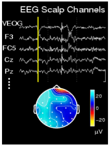
  Datasets are of different nature, resolution, and size

For example, consider data from three modalities

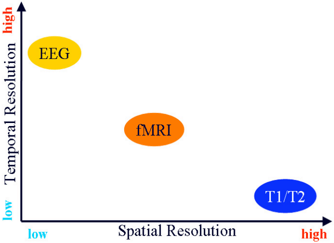
Structural MRI — *Morphology*



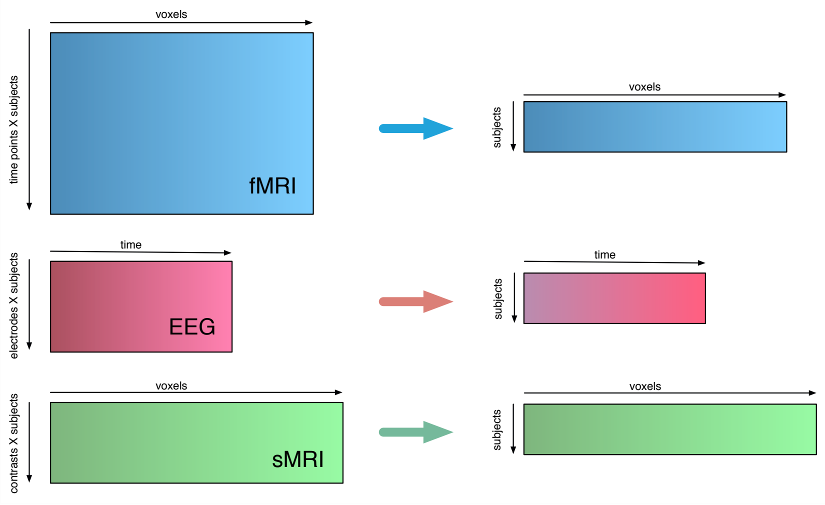
Functional MRI — *Function through blood oxygen level changes*



EEG — *Electrical field*

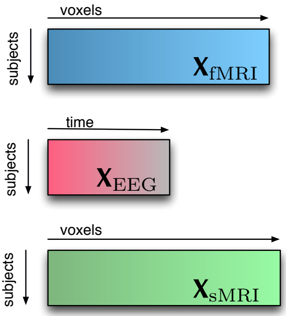


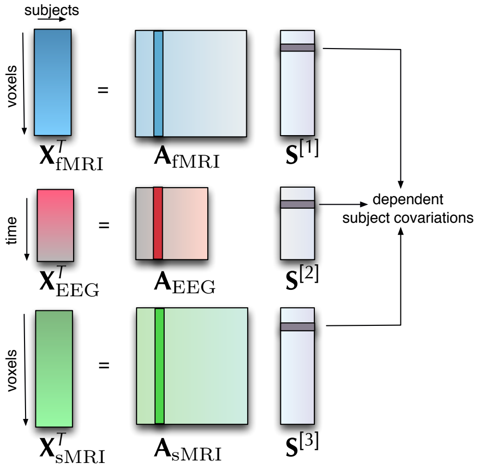
Extract *multivariate features* to create a dimension of coherence



Then need to identify the source of diversity, *the statistical link* across the datasets

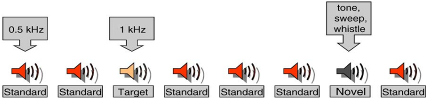
Still a little bit of  …



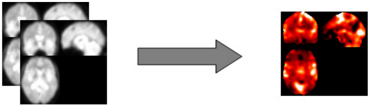


Transposed IVA (tIVA) model

Example using data collected during Auditory Oddball Task

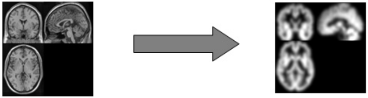


Features



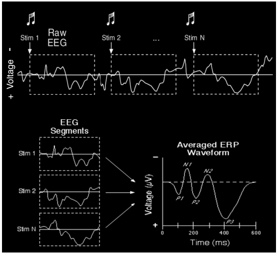
fMRI time series images

Task-related contrast images



sMRI images

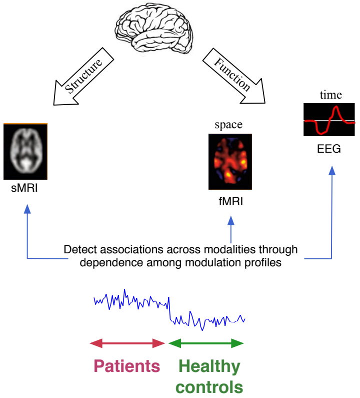
Segmented gray matter images

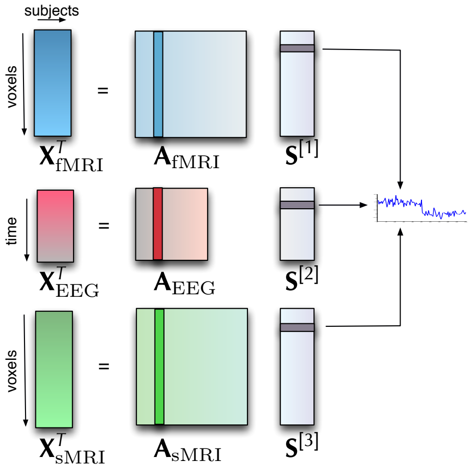


Event related potential (ERP)

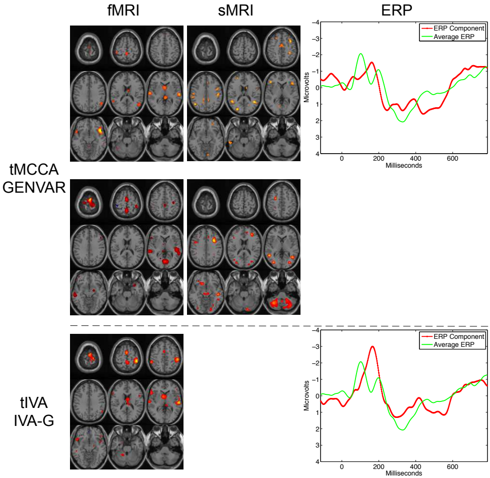
Data from 22 healthy controls and 14 patients with schizophrenia

A potential use: *Identify biomarkers through subject covariations*

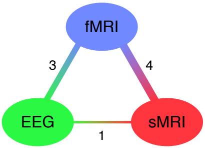




Patients with schizophrenia show less functional activity and less gray matter

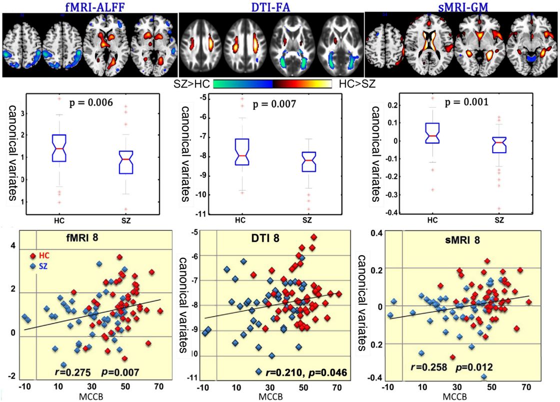


- Can identify components that discriminate only two modalities as well ($p<0.05$)
- Order selection enables exploratory analysis



[Adalı, *et al.*, 2015]

Using covariations, *can also study correlation with other variables*



47 schizophrenia patients (SZ) & 50 healthy controls (HC) & for red regions HC>SZ
Correlation with MATRICS Consensus Cognitive Battery

[Sui, *et al.*, 2015]

Summary

Joint use of multiple types of diversity enables

- identification of a broader class of signals
- maximal use of all available information
- and *true fusion* among multiple data sets

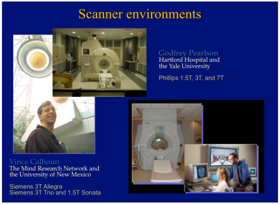
How "diverse" we would like to be?
Answer depends on many considerations

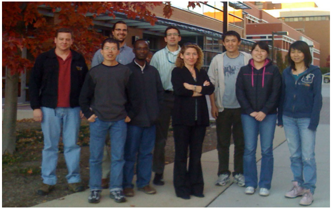
- model match
- computational cost
- robustness considerations
- *among many others...*

ICA, and more recently IVA, have proven fruitful for many applications, and there are many new possibilities for ICA, IVA, and beyond

Acknowledgments

MLSP-Lab
http://mlsp.umbc.edu

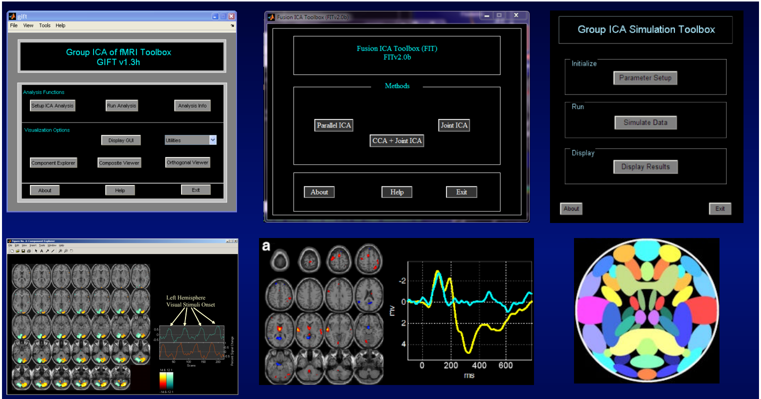




NSF Grants NSF-CCF 1117056 and NSF-IIS 1017718
NIH grants NIBIB R01 EB 005846 and NIBIB R01 EB 000840

Software packages

http://mialab.mrn.org/



GIFT
10000+ unique downloads

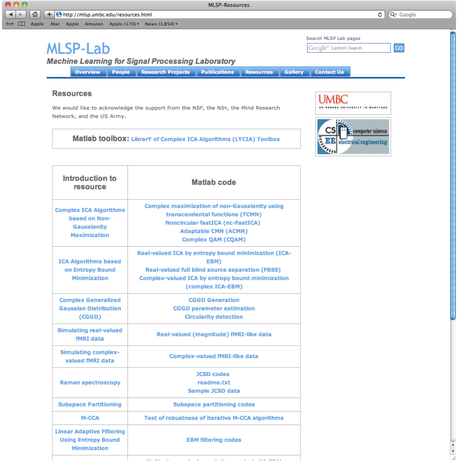
FIT
2000+ unique downloads

SimTB
600+ unique downloads

Funded by the NIH and the NSF

Matlab codes

http://mlsp.umbc.edu



MLSP-Resources

http://mlsp.umbc.edu/resources.html

Search MLSP Lab pages

MLSP-Lab

Machine Learning for Signal Processing Laboratory

Overview | People | Research Projects | Publications | Resources | Gallery | Contact Us

Resources

We would like to acknowledge the support from the NSF, the NIH, the Mind Research Network, and the US Army.

Matlab toolbox: LibraryY of Complex ICA Algorithms (LYCIA) Toolbox

Introduction to resource	Matlab code
Complex ICA Algorithms based on Non-Gaussianity Maximization	Complex maximization of non-Gaussianity using transcendental functions (TCMN) Noncircular fastICA (nc-FastICA) Adaptable CMN (ACMN) Complex QAM (CQAM)
ICA Algorithms based on Entropy Bound Minimization	Real-valued ICA by entropy bound minimization (ICA-EBM) Real-valued full blind source separation (FBSS) Complex-valued ICA by entropy bound minimization (complex ICA-EBM)
Complex Generalized Gaussian Distribution (CGGD)	CGGD Generation CGGD parameter estimation Circularity detection
Simulating real-valued fMRI data	Real-valued (magnitude) fMRI-like data
Simulating complex-valued fMRI data	Complex-valued fMRI-like data
Raman spectroscopy	JCSD codes readme.txt Sample JCSD data
Subspace Partitioning	Subspace partitioning codes
M-CCA	Test of robustness of iterative M-CCA algorithms
Linear Adaptive Filtering Using Entropy Bound Minimization	EBM filtering codes

Funded by the NIH and the NSF

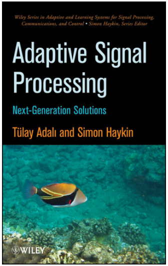
Few references, resources...

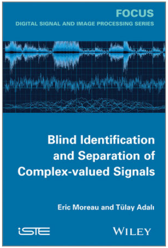
- *Review on ICA and IVA*, IEEE Signal Processing Magazine, May 2014, by Adalı, Anderson, and Fu

- *Special Issue on multi-modal data fusion*, Proceedings of the IEEE, September 2015, by Adalı, Jutten, and Hansen

On complex-valued signal processing and complex ICA:

- Blind Identification and Separation of Complex Signals by Moreau and Adalı, ISTE/Wiley 2013
- Adaptive Signal Processing: Next Generation Solutions by Adalı and Haykin, Wiley, 2010

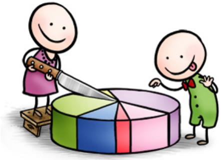




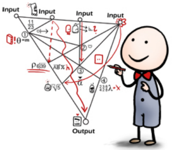
For references and additional resources: *http://mlsp.umbc.edu*

And a final note...

In many fields where data come from multiple sources, and is rich in structure, it pays off to be



data driven,



multivariate,

and not to forget to celebrate